Submitted, 2015 Chinese Control Conference (CCC)
http://www.cds.caltech.edu/~murray/papers/mm15-ccc.html

Multi-dimensional state estimation in adversarial environment

Yilin Mo[1], Richard M. Murray[1]
1. California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, United States.
yilinmo@caltech.edu, murray@cds.caltech.edu

Abstract: We consider the estimation of a vector state based on m measurements that can be potentially manipulated by an adversary. The attacker is assumed to have limited resources and can only manipulate up to l of the m measurements. However, it can the compromise measurements arbitrarily. The problem is formulated as a minimax optimization, where one seeks to construct an optimal estimator that minimizes the "worst-case" error against all possible manipulations by the attacker and all possible sensor noises. We show that if the system is not observable after removing $2l$ sensors, then the worst-case error is infinite, regardless of the estimation strategy. If the system remains observable after removing arbitrary set of $2l$ sensor, we prove that the optimal state estimation can be computed by solving a semidefinite programming problem. A numerical example is provided to illustrate the effectiveness of the proposed state estimator.

Key Words: Security, Estimation

1 Introduction

The increasing use of networked embedded sensors to monitor and control critical infrastructures provides potential malicious agents with the opportunity to disrupt their operations by corrupting sensor measurements. Supervisory Control And Data Acquisition (SCADA) systems, for example, run a wide range of safety critical plants and processes, including manufacturing, water and gas treatment and distribution, facility control, and power grids. A wide variety of motivations exists for launching an attack on the such systems, ranging from financial reasons, e.g., reducing the electricity bill, all the way to terrorism, e.g., threatening the life of possibly an entire population by controlling electricity and other life-critical resources. A successful attack to such kind of systems may significantly hamper the economy, the environment, and may even lead to the loss of human life. The first-ever SCADA system malware (called Stuxnet) was found in July 2010 and rose significant concern about SCADA system security [1, 2]. The research community has acknowledged the importance of addressing the challenge of designing secure detection, estimation and control systems [3].

We consider a secure estimation problem inspired by security concerns that arise from the possible manipulation of sensor data. We focus our attention on the estimation of a vector state $x \in \mathbb{R}^n$ from measurements collected by m sensors, with the caveat that the measurements are disturbed by an L_2 bounded noise and some of them can be further manipulated by a malicious third party. Limitations in the resources available to the attacker enable it to only manipulate l of the m sensors. However, the attacker has total control over the corrupted sensors, as it can change the measurements of the compromised sensors arbitrarily. To minimize the estimator's performance degradation in the presence of such attacks, we construct minimax estimator that minimize the "worst-case" expected cost against all possible noise and attacker's manipulate.

We show that if the system becomes unobservable after removing $2l$ sensor measurements, then even the optimal state estimator will have a "worst-case" unbounded error. For the case where the system remains observable after removing an arbitrary set of $2l$ sensors, we provide the explicit form of the optimal estimation, which is given by the Chebyshev center of a union of ellipsoids, which can be computed via semidefinite programming.

Related Work

Robust estimators such as M-estimator, L-estimator, R-estimator and etc. have also been extensively studied in the literature [4–6]. However, such approaches usually assume that the outliers of the data are generated *independently* by some other probability distribution different from the model assumptions. Furthermore, the robustness is usually measured by breakdown points [7, 8] or influence functions [9]. However, the independent assumptions do not hold in security settings. As the attacker can take control over multiple sensors, the compromised measurements from these sensors can be jointly selected by the adversary to maximize the estimation error. As a result, in this paper, we design our estimator to minimize the "worst-case" L_2 error against all possible attacks. Therefore, the concepts of robustness and security are different from each other. In other words, a robust estimator may not necessarily be secure and thus the techniques developed for robust estimation need to be re-examined before they can be applied in the context of security.

Furthermore, bad data detection and identification techniques, which are based on truncating the "atypical" data, have been widely used in large scale systems such as the power grid [10]. While such approaches are very successful in detecting and removing random failures, they are not effective against integrity attacks. Liu et al. [11] illustrate how an adversary can inject a stealthy input into the measurements to change the state estimate, without being detected by the bad data detector. Sandberg et al. [12] consider how to find a sparse stealthy input, which enables the adversary to launch an attack with a minimum number of compromised sensors. Xie et al. [13] further illustrate that the stealthy integrity attacks on state estimation can lead to a financial gain in the electricity market for the adversary.

For dynamical systems, a widely used approach is to construct "failure-sensitive" filters [14]. This detection scheme

This work is supported in part by IBM and UTC through iCyPhy consortium.

has been investigated recently in the context of cyber-physical security [15–18]. In these scenarios, the attacker can either arbitrarily perturb the system along certain directions without being detected by any filter or cannot induce any perturbation, without incurring detection. However, in the majority of these contributions, the system model is assumed to be noiseless, which greatly favors the failure detector, since the evolution of the system is deterministic and any deviation from the predetermined trajectory can be detected. A more realistic system model with bounded noise is considered by Pajic et al. [19]. They propose an estimator by solving an L_0 norm minimization problem and provide performance bound on the estimation error. In [20, 21], the authors also consider a noisy system, providing an algebraic condition under which an attacker can successfully destabilize the system and characterizing the performance of state estimators in this scenario.

This paper generalizes our previous works on secure estimation [22, 23], which consider designing the optimal estimator for a scalar state and minimizes the "worst-case" mean squared error. In this paper, we derive the optimal estimator for a vector state, with the caveat that the noise is L_2 bounded.

The rest of paper is organized as follows: In Section 2 we provide some preliminary results on the radius and diameter of compact set in Euclidean space. In Section 3 we formulate the problem of secure estimation with l manipulated measurements from m total measurements. In Section 4, we characterize the performance of the optimal estimator and provide an observability condition under which the estimator can have an unbounded error. In Section 5, we provide an algorithm to compute the optimal state estimate via semidefinite programming. In Section 6 we provide a numerical example to illustrate the proposed algorithm. Finally, Section 7 concludes the paper.

Notation

Let $x \in \mathbb{R}^n$ be a vector, then $\|x\|$ is the 2-norm of x. $\|x\|_0$ is the zero "norm" of x, i.e., the number of non-zero entries of x.

All comparisons between matrices are in the positive semidefinite sense.

2 Preliminary

A ball $B(x, r) \subset \mathbb{R}^n$ is defined as

$$B(x, r) \triangleq \{x' \in \mathbb{R}^n : \|x' - x\|_2 \leq r\}.$$

Consider a set $S \subseteq \mathbb{R}^n$. A ball $B(x, r)$ covers S if and only if $S \subseteq B(x, r)$. For any point $x \in \mathbb{R}^n$, define

$$\rho(x, S) \triangleq \inf\{r \in \mathbb{R}^+ : S \subseteq B(x, r)\}$$

We will assume that the infimum over an empty set is ∞. Hence, if S is unbounded, then $\rho(x, S) = \infty$ for any x.

For a bounded set S, define the radius $r(S) \in \mathbb{R}^+$ and Chebyshev center $c(S) \in \mathbb{R}^n$ of S to be

$$r(S) \triangleq \inf_{x \in \mathbb{R}^n} \rho(x, S),$$
$$c(S) \triangleq \arg\min_{x \in \mathbb{R}^n} \rho(x, S).$$

In an essence, $B(c(S), r(S))$ is the smallest radius ball that covers S. Notice that $c(S)$ may *not* necessarily belong to S. For an unbounded S, we define $r(S) = \infty$.

We further define the diameter $d(S)$ of the set S as

$$d(S) \triangleq \sup_{x \in S} \rho(x, S).$$

Notice that in general $d(S) \neq 2r(S)$. For example, for a equilateral triangle with side length 1, we have $d(S) = 1$, while $r(S) = 1/\sqrt{3}$. In general, the following relation between $r(S)$ and $d(S)$ holds:

Theorem 1. *Let $S \subset R^n$ be a non-empty and bounded set, then the following inequalities hold on $r(S)$ and $d(S)$*

$$\frac{d(S)}{2} \leq r(S) \leq \sqrt{\frac{n}{2n+2}} d(S) \leq \frac{1}{\sqrt{2}} d(S).$$

Proof. The first inequality is due to the fact that $S \subseteq B(c(S), r(S))$, which implies that

$$d(S) \leq d[B(c(S), r(S))] = 2r(S).$$

The second inequality is from Jung's theorem [24]. The third inequality is trivial. □

3 Problem Formulation

The goal is to estimate the state $x \in \mathbb{R}^n$ from a vector $y \triangleq [y_1, \ldots, y_m]^T \in \mathbb{R}^m$ consisting of m sensor measurements $y_i \in \mathbb{R}$, where the index $i \in \mathcal{S} \triangleq \{1, 2, ..., m\}$. The measurements could potentially be compromised by an adversary. Therefore, we assume that x and y satisfies the following equation:

$$y = Hx + Gw + a, \tag{1}$$

where $\|w\| \leq \delta$ is the sensor noise, which is assumed to be bounded, and $G \in \mathbb{R}^{m \times m}$ is assumed to be full rank. The vector a is the bias injected by the attacker. The non-zero entries of a indicates the set of compromised sensors. In this paper, we assume that the attacker can only manipulate up to l sensors. As a result, $\|a\|_0 \leq l$.

Remark 1. The parameter l can also be interpreted as a design parameter for the system operator. In general, increasing l will increase the resilience of the estimator under attack. However, a large l could result in performance degradation during normal operation when no sensor is compromised. Therefore, there exists a trade-off between resilience and efficiency (under normal operation), which can be tuned by choosing a suitable parameter l.

Define

$$H \triangleq \begin{bmatrix} h_1 \\ \vdots \\ h_m \end{bmatrix}, w \triangleq \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}, a \triangleq \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix},$$

and define the set $\mathbb{Y}$ as the set of all possible "manipulated" measurements, i.e.,

$$\mathbb{Y} \triangleq \{y \in \mathbb{R}^m : \exists\, x,\, w,\, a,$$
$$\text{such that } \|w\|_2 \leq \delta,\, \|a\|_0 \leq l \text{ and } y = Hx + Gw + a\}.$$

For any $y \in \mathbb{Y}$, we can define the set $\mathbb{X}(y)$ as the set of feasible x that can generate y, i.e.,

$$\mathbb{X}(y) \triangleq \{x \in \mathbb{R}^n : \exists\, w,\, a,\\ \text{such that } \|w\| \le \delta,\, \|a\|_0 \le l \text{ and } y = Hx + Gw + a\}.$$

An estimator is a function $f : \mathbb{Y} \to \mathbb{R}^n$, where $\hat{x} = f(y)$. Given y, the magnitude of the worst case estimation error is defined as

$$e(y) \triangleq \sup_{x \in \mathbb{X}(y)} \|f(y) - x\|.$$

From the definition of the Chebyshev center, we know that the optimal estimator with smallest worst case error e is given by

$$f^*(y) \triangleq c(\mathbb{X}(y)),$$

with worst case error

$$e(y) = r(\mathbb{X}(y)).$$

Therefore, the worst case error magnitude for all possible y is given by

$$e^* \triangleq \sup_{y \in \mathbb{Y}} r(\mathbb{X}(y)).$$

In the following sections, we provides an upper and lower bound for e^*. We further propose an algorithm to compute $c(\mathbb{X}(y))$ via convex optimization.

4 Performance Bounds for the Optimal Estimator

This section is devoted to analyzing the performance of the optimal estimator. To this end, for any index set $\mathcal{I} = \{i_1, \ldots, i_j\}$, define the complement set $\mathcal{I}^c = \mathcal{S} \backslash I$ and define subspace $\mathcal{V}_\mathcal{I} \triangleq \text{span}(e_{i_1}, \ldots, e_{i_j}) \subseteq \mathbb{R}^m$, where $e_i \in \mathbb{R}^m$ is the ith canonical basis vector. Define the following set:

$$\mathbb{X}_\mathcal{I}(y) \triangleq \{x \in \mathbb{R}^n : \exists\, w,\, a \in \mathcal{V}_{\mathcal{I}^c},\\ \text{such that } \|w\|_2 \le \delta \text{ and } y = Hx + Gw + a\}.$$

Hence, $\mathbb{X}_\mathcal{I}(y)$ represents all possible states that can generate measurement y when the sensors in $\mathcal{I}$ are good and the sensors in $\mathcal{I}^c$ are compromised. By enumerating all possible $\mathcal{I}$s, it is easy to see that $\mathbb{X}(y)$ can be written as

$$\mathbb{X}(y) = \bigcup_{|\mathcal{I}| = m-l} \mathbb{X}_\mathcal{I}(y).$$

For any $\mathcal{I} = \{i_1, \ldots, i_j\}$, we can define

$$H_\mathcal{I} \triangleq \begin{bmatrix} h_{i_1} \\ \vdots \\ h_{i_j} \end{bmatrix}, G_\mathcal{I} \triangleq \begin{bmatrix} g_{i_1} \\ \vdots \\ g_{i_j} \end{bmatrix}, y_\mathcal{I} \triangleq \begin{bmatrix} y_{i_1} \\ \vdots \\ y_{i_j} \end{bmatrix},$$

where g_i is the ith row vector of G.

$$F_\mathcal{I} \triangleq G_\mathcal{I} G_\mathcal{I}^T.$$

Since G is full rank, $G_\mathcal{I}$ is full row rank, which implies that $F_\mathcal{I}$ is full rank. Thus, if $H_\mathcal{I}$ is full column rank, we can define

$$K_\mathcal{I} \triangleq \left(H_\mathcal{I}^T F_\mathcal{I}^{-1} H_\mathcal{I}\right)^{-1} H_\mathcal{I}^T F_\mathcal{I}^{-1},$$
$$P_\mathcal{I} \triangleq \left(H_\mathcal{I}^T F_\mathcal{I}^{-1} H_\mathcal{I}\right)^{-1},$$
$$U_\mathcal{I} \triangleq (I - H_\mathcal{I} K_\mathcal{I})^T F_\mathcal{I}^{-1} (I - H_\mathcal{I} K_\mathcal{I}).$$

The following theorem provides bounds on e^*:

Theorem 2. *If there exists an index set $\mathcal{K} \subset \mathcal{S}$ with cardinality $m - 2l$, such that $H_\mathcal{K}$ is not of full column rank, then $e^* = \infty$. If for all $|\mathcal{K}| = m - 2l$, $H_\mathcal{K}$ is full column rank, then for all possible $y \in \mathbb{Y}$, we have*

$$\sup_{y \in \mathbb{Y}} d(\mathbb{X}(y)) = 2\delta \max_{|\mathcal{K}| = m-2l} \sqrt{\sigma(P_\mathcal{K})}. \tag{2}$$

Therefore, e^* satisfies

$$\max_{|\mathcal{K}| = m-2l} \delta\sqrt{\sigma(P_\mathcal{K})} \le e^* \le \max_{|\mathcal{K}| = m-2l} \delta\sqrt{2\sigma(P_\mathcal{K})}, \tag{3}$$

where $\sigma(P)$ is the spectral radius of P.

Before proving Theorem 2, we need the following lemma:

Lemma 1. *If $\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \mathcal{S}$ and $H_{\mathcal{K}_1}$ is full column rank, then the following statement holds:*

1) *$H_{\mathcal{K}_2}$ is also full column rank.*
2) *$P_{\mathcal{K}_1} \ge P_{\mathcal{K}_2}$.*

Proof. Without loss of generality, let us assume that

$$H_{\mathcal{K}_2} = \begin{bmatrix} H_{\mathcal{K}_1} \\ H_{\mathcal{K}_2} \end{bmatrix}. \tag{4}$$

Therefore,

$$n \ge \text{rank}(H_{\mathcal{K}_2}) \ge \text{rank}(H_{\mathcal{K}_1}) = n.$$

Hence, $H_{\mathcal{K}_2}$ is full column rank, which implies that $P_{\mathcal{K}_2}$ is well-defined.

To prove $P_{\mathcal{K}_1} \ge P_{\mathcal{K}_2}$, we only need to prove that

$$H_{\mathcal{K}_1}^T F_{\mathcal{K}_1}^{-1} H_{\mathcal{K}_1} \le H_{\mathcal{K}_2}^T F_{\mathcal{K}_2}^{-1} H_{\mathcal{K}_2}. \tag{5}$$

From definition of $F_\mathcal{I}$, we can write $F_{\mathcal{K}_2}$ as

$$F_{\mathcal{K}_2} = \begin{bmatrix} F_{11} & F_{12} \\ F_{12}' & F_{22}, \end{bmatrix}$$

where $F_{11} = F_{\mathcal{K}_1}$. Using Schur complements, we have

$$F_{\mathcal{K}_2}^{-1} = \begin{bmatrix} F_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} F_{11}^{-1} F_{12} \\ I \end{bmatrix} \left(F_{22} - F_{12}' F_{11}^{-1} F_{12}\right)^{-1} \begin{bmatrix} F_{12}' F_{11}^{-1} & I \end{bmatrix}$$

Combining with (4), we can prove (5). □

We are now ready to prove Theorem 2:

Proof of Theorem 2. We first prove (2). Suppose that for all $\mathcal{K} \subset \mathcal{S}$ with cardinality $m - 2l$, $H_\mathcal{K}$ is full column rank. Let us consider a pair of set $\mathcal{I}$, $\mathcal{J}$ with cardinality $m - l$. Define $\mathcal{K}$ as

$$\mathcal{K} = \mathcal{I} \bigcap \mathcal{J} = \mathcal{S} \backslash \left(\mathcal{I}^c \bigcup \mathcal{J}^c\right)$$

Clearly, $|\mathcal{K}| \ge m - 2l$ and it includes a index set of size $m - 2l$. Hence, by Lemma 1, $H_\mathcal{K}$ is also full column rank. Now for any point $x_1 \in \mathbb{X}_\mathcal{I}(y)$ and $x_2 \in \mathbb{X}_\mathcal{J}(y)$, we have:

$$Hx_1 + Gw_1 + a_1 = Hx_2 + Gw_2 + a_2 = y, \tag{6}$$

where $a_1 \in \mathcal{V}_{\mathcal{I}^c}$ and $a_2 \in \mathcal{V}_{\mathcal{J}^c}$. Since both a_1 and a_2 have zero entries on the ith entry, where $i \in \mathcal{K}$, (6) implies that

$$H_\mathcal{K} x_1 + G_\mathcal{K} w_1) = H_\mathcal{K} x_2 + G_\mathcal{K} w_2). \tag{7}$$

which implies that

$$x_1 - x_2 = K_{\mathcal{K}} G_{\mathcal{K}} (w_2 - w_1). \tag{8}$$

By the fact that $\|w_2 - w_1\|_2 \leq 2\delta$, we have

$$\|x_1 - x_2\| \leq 2\|K_{\mathcal{K}} G_{\mathcal{K}}\|\delta = 2\delta\sqrt{\sigma(P_{\mathcal{K}})},$$

where $\|K_{\mathcal{K}} G_{\mathcal{K}}\|$ is the largest singular value of $K_{\mathcal{K}} G_{\mathcal{K}}$. Therefore, by Lemma 1, we have

$$\begin{aligned}\sup_{y\in\mathbb{Y}} d(\mathbb{X}(y)) &\leq 2\delta \max_{|\mathcal{K}|\geq m-2l} \sqrt{\sigma(P_{\mathcal{K}})} \\ &= 2\delta \max_{|\mathcal{K}|=m-2l} \sqrt{\sigma(P_{\mathcal{K}})}.\end{aligned}$$

Now we need to prove that the equality of (2) holds. Suppose that we find x_1, x_2 and $\|w_1\|, \|w_2\| \leq \delta$ that satisfies (7) and $\|x_1 - x_2\| = 2\sqrt{\sigma(P_{\mathcal{K}})}$. We know that

$$H_{\mathcal{K}} x_1 + G_{\mathcal{K}} w_1 = H_{\mathcal{K}} x_2 + G_{\mathcal{K}} w_2. \tag{9}$$

Therefore, let us create a y, such that

$$\begin{aligned} y_{\mathcal{K}} &= H_{\mathcal{K}} x_1 + G_{\mathcal{K}} w_1, \\ y_{\mathcal{I}\setminus\mathcal{K}} &= H_{\mathcal{I}\setminus\mathcal{K}} x_1 + G_{\mathcal{I}\setminus\mathcal{K}} w_1, \\ y_{\mathcal{J}\setminus\mathcal{K}} &= H_{\mathcal{J}\setminus\mathcal{K}} x_2 + G_{\mathcal{J}\setminus\mathcal{K}} w_2, \\ y_{\mathcal{S}\setminus(\mathcal{I}\cup\mathcal{J})} &= 0.\end{aligned}$$

Thus,

$$\begin{aligned} y_{\mathcal{K}} - H_{\mathcal{K}} x_1 &= G_{\mathcal{K}} w_1, \\ y_{\mathcal{I}\setminus\mathcal{K}} - H_{\mathcal{I}\setminus\mathcal{K}} x_1 &= G_{\mathcal{I}\setminus\mathcal{K}} w_1,\end{aligned}$$

which implies that $x_1 \in \mathbb{X}_{\mathcal{I}}(y)$. On the other hand,

$$\begin{aligned} y_{\mathcal{K}} - H_{\mathcal{K}} x_2 = H_{\mathcal{K}}(x_1 - x_2) + G_{\mathcal{K}} w_1 = G_{\mathcal{K}} w_2, \\ y_{\mathcal{J}\setminus\mathcal{K}} - H_{\mathcal{J}\setminus\mathcal{K}} x_2 = G_{\mathcal{J}\setminus\mathcal{K}} w_2,\end{aligned}$$

which implies that $x_2 \in \mathbb{X}_{\mathcal{J}}(y)$. Therefore, (2) holds. (3) can be proved by applying Theorem 1.

Now suppose there exists an $|\mathcal{K}| = m - 2l$, such that $H_{\mathcal{K}}$ is not full column rank. We can find index set $\mathcal{I}, \mathcal{J}$, such that $|\mathcal{I}| = |\mathcal{J}| = m - l$ and $\mathcal{I} \bigcap \mathcal{J} = \mathcal{K}$. Furthermore, we know that there exists $x_1 \neq 0$, such that

$$H_{\mathcal{K}} x_1 = 0.$$

As a result, if we choose $x_2 = 0$, $w_1 = w_2 = 0$, then (9) holds for x_1, x_2, w_1, w_2. Now by the similar argument, we can construct a y, such that $x_1 \in \mathbb{X}_{\mathcal{I}}(y)$ and $x_2 = 0 \in \mathbb{X}_{\mathcal{J}}(y)$. Moreover, by linearity, we know that

$$\alpha x_1 \in \mathbb{X}_{\mathcal{I}}(\alpha y),\ 0 = \alpha x_2 \in \mathbb{X}_{\mathcal{J}}(\alpha y).$$

Hence, by Theorem 1,

$$e^* \geq \sup_{y\in\mathbb{Y}} \frac{d(\mathbb{X}(y))}{2} \geq \sup_{\alpha\in\mathbb{R}} \frac{\|\alpha x_1\|}{2} = \infty.$$

□

5 Estimator Design

In this section, we first characterize the shape of $\mathbb{X}_{\mathcal{I}}(y)$:

Theorem 3. *Define the function $V_{\mathcal{I}}(x) : \mathbb{R}^n \to \mathbb{R}$ as the solution of the following optimization problem:*

$$\begin{aligned} &\underset{w\in\mathbb{R}^m}{\textit{minimize}} && \|w\|^2 \\ &\textit{subject to} && G_{\mathcal{I}} w = y_{\mathcal{I}} - H_{\mathcal{I}} x. \end{aligned} \tag{10}$$

Then $V_{\mathcal{I}}(x)$ is given by

$$V_{\mathcal{I}}(x) = (x - \hat{x}_{\mathcal{I}}(y))^T P_{\mathcal{I}}^{-1} (x - \hat{x}_{\mathcal{I}}(y)) + \varepsilon_{\mathcal{I}}(y), \tag{11}$$

where

$$\hat{x}_{\mathcal{I}}(y) = K_{\mathcal{I}} y_{\mathcal{I}}, \tag{12}$$

and

$$\varepsilon_{\mathcal{I}}(y) = y_{\mathcal{I}}^T U_{\mathcal{I}} y_{\mathcal{I}}. \tag{13}$$

Proof. Consider the constraint of the optimization problem (10)

$$y_{\mathcal{I}} - H_{\mathcal{I}} x = G_{\mathcal{I}} w. \tag{14}$$

As G is full row rank, $G_{\mathcal{I}}$ is also full row rank. Consider the singular value decomposition of $G_{\mathcal{I}}$, we get

$$G_{\mathcal{I}} = Q_1 \begin{bmatrix}\Lambda & \mathbf{0}\end{bmatrix} Q_2,$$

where Q_1, Q_2 are orthogonal matrices with proper dimensions and Λ is an invertible and diagonal matrix. Hence, (14) implies that

$$\Lambda^{-1} Q_1^T y_{\mathcal{I}} - \Lambda^{-1} Q_1^T H_{\mathcal{I}} x = \begin{bmatrix} I & \mathbf{0}\end{bmatrix} v. \tag{15}$$

where $v = Q_2 w$ and $\|v\| = \|w\|$. By projecting $\Lambda^{-1} Q_1^T y_{\mathcal{I}}$ into the subspace span$(\Lambda^{-1} Q_1^T H_{\mathcal{I}})$, we have

$$\begin{aligned}\left[\Lambda^{-1} Q_1^T y_{\mathcal{I}} - \Lambda^{-1} Q_1^T H_{\mathcal{I}} \hat{x}_{\mathcal{I}}(y)\right] + \Lambda^{-1} Q_1^T H_{\mathcal{I}} \left[x - \hat{x}_{\mathcal{I}}(y)\right] \\ = \begin{bmatrix} I & \mathbf{0}\end{bmatrix} v.\end{aligned} \tag{16}$$

The first term on the LHS of (16) is perpendicular to the second term. Thus, (16) is equivalent to

$$\begin{aligned}\varepsilon_{\mathcal{I}}(y) + (x - \hat{x}_{\mathcal{I}}(y))^T P_{\mathcal{I}}^{-1} (x - \hat{x}_{\mathcal{I}}(y)) \\ = \left\| \begin{bmatrix} I & \mathbf{0}\end{bmatrix} v \right\|^2 \leq \|v\|^2 = \|w\|^2.\end{aligned}$$

Clearly, the equality holds when $v = \begin{bmatrix} v_1, \ldots, v_{|\mathcal{I}|}, 0, \ldots, 0\end{bmatrix}$. Hence

$$V_{\mathcal{I}}(x) = \varepsilon_{\mathcal{I}}(y) + (x - \hat{x}_{\mathcal{I}}(y))^T P_{\mathcal{I}}^{-1} (x - \hat{x}_{\mathcal{I}}(y)).$$

□

By Theorem 3, we immediately have the following corollary:

Corollary 1. *If $\varepsilon_{\mathcal{I}}(y) > \delta^2$, then $\mathbb{X}_{\mathcal{I}}(y)$ is an empty set. Otherwise, $\mathbb{X}_{\mathcal{I}}(y)$ is an ellipsoid given by*

$$\mathbb{X}_{\mathcal{I}}(y) = \{x : (x - \hat{x}_{\mathcal{I}}(y))^T P_{\mathcal{I}}^{-1} (x - \hat{x}_{\mathcal{I}}(y)) \leq \delta^2 - \varepsilon_{\mathcal{I}}(y)\}. \tag{17}$$

Proof. By definition, $x \in \mathbb{X}_{\mathcal{I}}(y)$ is equivalent to the existence of w, such that $\|w\| \leq \delta$ and

$$y_{\mathcal{I}} = H_{\mathcal{I}} x + G_{\mathcal{I}} w.$$

Hence, the corollary holds by Theorem 3. □

Remark 2. One can view $\varepsilon_{\mathcal{I}}(y)$ as the deviation of the measurement from the attack model. If $\varepsilon_{\mathcal{I}}(y) \geq \delta^2$, i.e., the deviation cannot be explained by the noise, then $\mathbb{X}_{\mathcal{I}}(y)$ is empty, which implies that the good sensor set is not $\mathcal{I}$.

Let us define set as

$$\mathfrak{I} \triangleq \{\mathcal{I} \subset \mathcal{S} : |\mathcal{I}| = m - l,\, \delta^2 \geq \varepsilon_{\mathcal{I}}(y)\}. \tag{18}$$

Since $\mathbb{X}(y) = \bigcup_{|\mathcal{I}|=l} \mathbb{X}_{\mathcal{I}}(y) = \bigcup_{\mathcal{I}\in\mathfrak{I}} \mathbb{X}_{\mathcal{I}}(y)$, we know that $\mathbb{X}(y)$ is a union of ellipsoids. To check if a ball covers a union of ellipsoids, we have the following theorem [25]:

Theorem 4. *A ball $B(x, r)$ covers $\mathbb{X}(y)$ if and only if for every index set $|\mathcal{I}| = m - l$, such that*

$$\delta^2 - \varepsilon_{\mathcal{I}}(y) \geq 0,$$

there exists $\tau_{\mathcal{I}} \geq 0$, such that

$$\tau_{\mathcal{I}} \Omega_{\mathcal{I}} \geq \begin{bmatrix} I & -x & 0 \\ -x^T & r^2 & x^T \\ 0 & x & -I \end{bmatrix}, \tag{19}$$

where $\Omega_{\mathcal{I}}$ is defined as,

$$\Omega_{\mathcal{I}} = \begin{bmatrix} P_{\mathcal{I}}^{-1} & -P_{\mathcal{I}}^{-1}\hat{x}_{\mathcal{I}}(y) & 0 \\ -\hat{x}_{\mathcal{I}}(y)^T P_{\mathcal{I}}^{-1} & \hat{x}_{\mathcal{I}}(y)^T P_{\mathcal{I}}^{-1}\hat{x}_{\mathcal{I}}(y) + \varepsilon_{\mathcal{I}}(y) - \delta^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Proof. This theorem can be proved by Lemma 2.8 in [25]. □

Therefore, we can derive the optimal state estimate as the solution of the following semidefinite programming problem:

$$\begin{aligned} \underset{\hat{x},\varphi,\tau_{\mathcal{I}}}{\text{minimize}} \quad & \varphi \\ \text{subject to} \quad & \varphi \geq 0, \\ & \tau_{\mathcal{I}} \geq 0,\, \forall \mathcal{I} \in \mathfrak{I}, \\ & \tau_{\mathcal{I}} \Omega_{\mathcal{I}} \geq \begin{bmatrix} I & -\hat{x} & 0 \\ -\hat{x}^T & \varphi & \hat{x}^T \\ 0 & \hat{x} & -I \end{bmatrix}, \forall \mathcal{I} \in \mathfrak{I}. \end{aligned} \tag{20}$$

where the radius of the Chebyshev ball is $r = \sqrt{\varphi}$.

In conclusion, the optimal state estimation can be computed via the following algorithm:

1) Enumerate all possible $|\mathcal{I}| = m-l$, compute $\hat{x}_{\mathcal{I}}(y)$ and $\varepsilon_{\mathcal{I}}(y)$ via (12) and (13).
2) Check whether $\varepsilon_{\mathcal{I}}(y)$ is no greater than δ^2. Compute the index set $\mathfrak{I}$ via (18).
3) Solve the optimization problem (10).

6 Numerical Example

In this section, we provide a numerical example to illustrate our estimator design. We assume that $n = 2$, $m = 4$ and one sensor is compromised. The noise is assume to satisfy $\|w\| \leq \delta = 1$. We further assume that

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}, G = I.$$

It is easy to check that for any $|K| = m - 2l = 2$, $H_{\mathcal{K}}$ is full column rank.

We first consider the worst case performance of our estimator. One can verify that

$$\max_{|\mathcal{K}|=2} \sigma(P_{\mathcal{K}}) = 2.618$$

where the corresponding $\mathcal{K} = \{1, 3\}$. Using the procedure described in the proof of Theorem 2, we choose $\mathcal{I} = \{1, 2, 3\}$, $\mathcal{J} = \{1, 3, 4\}$. We can then construct the following variables:

$$x_1 = \begin{bmatrix} -1.701 \\ 2.753 \end{bmatrix},\, w_1 = \begin{bmatrix} 0.8507 \\ 0 \\ -0.5257 \\ 0 \end{bmatrix}.$$

and

$$x_2 = 0,\, w_2 = \begin{bmatrix} -0.8507 \\ 0 \\ 0.5257 \\ 0 \end{bmatrix}.$$

The corresponding y is given by

$$y = \begin{bmatrix} -0.851 \\ 2.753 \\ 0.5257 \\ 0 \end{bmatrix}.$$

The optimal state estimate $\hat{x}$ and worst-case error $e(y)$ is given by

$$\hat{x} = \begin{bmatrix} -0.851 \\ 1.376 \end{bmatrix},\, e(y) = 1.618.$$

On the other hand, if the system operator is unaware of the existence of the adversary, it is easy to prove that the optimal estimator designed for the non-adversarial environment is given by

$$\hat{x} = \left(H^T (GG^T)^{-1} H\right)^{-1} H^T (GG^T)^{-1} y. \tag{21}$$

For our case, the state estimate given by (21) is

$$\hat{x} = \begin{bmatrix} -0.108 \\ 1.093 \end{bmatrix},$$

for which the worst case error $e(y) = 2.306$.

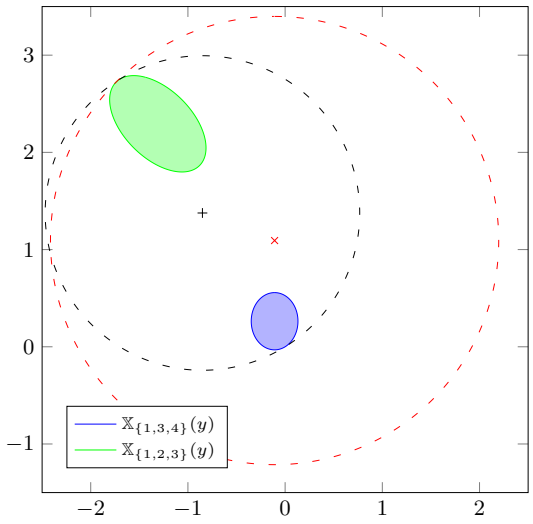

Fig. 1: The performance of the optimal state estimator. The green ellipse corresponds to $\mathbb{X}_{\{1,3,4\}}(y)$ and the red ellipse corresponds to $\mathbb{X}_{\{1,2,3\}}(y)$. The set $\mathbb{X}_{\{2,3,4\}}(y)$ and $\mathbb{X}_{\{1,2,4\}}(y)$ is empty. The black "+" is the optimal state estimate while the black dashed line is the Chebyshev ball for $\mathbb{X}(y)$. The red "$\times$" corresponds to the output of the optimal state estimator designed for benign sensors and the red dashed line is the minimum covering ball of $\mathbb{X}(y)$ centered at "$\times$".

7 Conclusion and Future Work

We consider the estimation of a vector state x based on m measurements, where l of them are malicious and can be changed arbitrarily by an adversary. We prove that if the system is not observable after removing $2l$ sensor measurements, then the attacker can make the worst case estimation error to be infinite. On the other hand, we provides upper and lower bound for the worst case estimation error when the system remains observable after removing any set of $2l$ sensors. We then derive the optimal state estimation as the solution of a semidefinite programming problem.

In the future, we want to find a near optimal state estimator with lower computation complexity. Furthermore, we would like to consider stochastic noise models.

References

[1] T. M. Chen, "Stuxnet, the real start of cyber warfare? [editor's note]," vol. 24, no. 6, pp. 2–3, 2010.

[2] D. P. Fidler, "Was stuxnet an act of war? decoding a cyber-attack," *IEEE Security & Privacy*, vol. 9, no. 4, pp. 56–59, 2011.

[3] A. A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems," in *HOTSEC'08: Proceedings of the 3rd conference on Hot topics in security*. Berkeley, CA, USA: USENIX Association, 2008, pp. 1–6.

[4] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," vol. 73, no. 3, pp. 433–481, 1985.

[5] R. A. Maronna, D. R. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. Wiley, 2006.

[6] P. J. Huber and E. M. Ronchetti, *Robust Statistics*. Wiely, 2009.

[7] F. R. Hampel, "A general qualitative definition of robustness," *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, Dec 1971.

[8] D. L. Donoho and P. J. Huber, "The notion of breakdown point," *A Festschrift for Erich L. Lehmann*, pp. 157–184, 1983.

[9] F. R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, 1974.

[10] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[11] Y. Liu, M. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM conference on Computer and communications security*, 2009.

[12] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems*, 2010.

[13] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[14] A. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, Nov 1976.

[15] F. Pasqualetti, A. Bicchi, and F. Bullo, "Consensus computation in unreliable networks: A system theoretic approach," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 90–104, Jan 2010.

[16] F. Pasqualetti, F. Dorfler, and F. Bullo, "Cyber-physical attacks in power networks: models, fundamental limitations and monitor design," in *Proc. 50th IEEE Conf. Decision and Control and European Control Conf. (CDC-ECC)*, 2011, pp. 2195–2201.

[17] S. Sundaram, M. Pajic, C. Hadjicostis, R. Mangharam, and G. J. Pappas, "The wireless control network: monitoring for malicious behavior," in *IEEE Conference on Decision and Contro*, Atlanta, GA, Dec 2010.

[18] H. Fawzi, P. Tabuada, and S. Diggavi, "Security for control systems under sensor and actuator attacks," in *Proc. IEEE 51st Annual Conf. Decision and Control (CDC)*, Maui, HI, 2012, pp. 3412–3417.

[19] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. Pappas, "Robustness of attack-resilient state estimators," in *Cyber-Physical Systems (ICCPS), 2014 ACM/IEEE International Conference on*, April 2014, pp. 163–174.

[20] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Proc. 49th IEEE Conf. Decision and Control (CDC)*, Atlanta, Georgia, 2010, pp. 5967–5972.

[21] Y. Mo and B. Sinopoli, "False data injection attacks in cyber physical systems," in *First Workshop on Secure Control Systems*, Stockholm, Sweden, April 2010.

[22] ——, "Robust estimation in the presence of integrity attacks," in *52nd IEEE Conference on Decision and Control*, 2013, pp. 6085 – 6090.

[23] ——, "Secure estimation in the presence of integrity attacks," *Automatic Control, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[24] L. Danzer, B. Grünbaum, and V. Klee, *Helly's Theorem and Its Relatives*, ser. Proceedings of symposia in pure mathematics: Convexity. American Mathematical Society, 1963.

[25] E. Yildirim, "On the minimum volume covering ellipsoid of ellipsoids," *SIAM Journal on Optimization*, vol. 17, no. 3, pp. 621–641, 2006. [Online]. Available: http://dx.doi.org/10.1137/050622560